WHAT IS CLAIMED IS:

1. A method of extracting data from a source document wherein words, portions of words and graphic elements are circumscribed by polygons hereinafter referred to as quads, comprising the steps of:

oversizing each quad in a predetermined range of the document;

selecting a first quad;

a. determining if said selected first quad intersects and thus overlaps a second quad:

i. if overlap is detected, assigning said second quad to the selected first quad, thereby creating a total polygon area larger than either original entity; and

ii. if overlap is not detected, creating an encompassing polygon including said selected first quad;

b. merging the newly created or enlarged polygon (of step "a") with any other polygon found to overlap with the step "a" polygon;

repeating steps "a" and "b" for any further unassigned quads left in said predetermined range and not encompassed by a polygon;

selecting a first frame of quads assigned to a polygon;

c. decomposing all graphical elements within the selected frame into straight lines;

d. assembling components of the selected frame into a table when examination of the decomposed frame indicates a predetermined arrangement of table cell defining straight lines; and

repeating steps "c" and "d" for identifying any possible table in any remaining frames of assigned quads.

2. The method of claim 1 wherein:

step c includes retaining relative placement data of each decomposed line as obtained from the source document; and

step d comprises the additional sub-steps of:

combining co-aligned lines in the frame;

creating a rectangular table boundary;

splitting any previously formed rectangle determined to be cut by any selected co-aligned line inserted into said rectangular table boundary to create at least two new rectangles; and

removing any created rectangle whose interior would not include at least one text quad.

3. A method of mining data from a visually displayable source document having quads defining textual and graphics elements, comprising the steps of:

combining quads, having less than a predetermined separation, into frames;

generating textual paragraphs, in an output document, from frames detected to contain only textual and textual associated symbols; and

generating tables, in said output document, from frames detected to contain only vertically and horizontally oriented straight lines enclosing textual and textual associated symbols.

4. A method of eliminating a group of quads in a source document, each quad of which is relatively situated within a predetermined distance of each other, as potentially defining a table of data, comprising the steps of:

decomposing all graphic elements in said group of quads into straight lines;

determining the orientation of each of said straight lines; and

eliminating said group of quads where any of said straight lines is oriented other than horizontal and vertical.

5. A method of eliminating a group of quads in a source document, each quad of which is relatively situated within a predetermined distance of each other, as potentially defining a table of data, comprising the steps of:

decomposing all graphic elements in said group of quads into straight lines;

combining any decomposed co-aligned lines into a new set of straight lines;

creating a rectangular table perimeter;

cutting said rectangular table perimeter into smaller rectangles using said straight lines whereby a plurality of cells are created;

removing rectangles not circumscribing a text quad of said group of quads;

eliminating said group of quads, as potentially defining a table, if any text quad is not circumscribed by a rectangular cell.

6. A method of establishing that a group of quads in a source document, each of which is relatively situated within a predetermined distance of each other, potentially defines a table of data, comprising the steps of:

decomposing all graphic elements in said group of quads into straight lines; and

determining that the orientation of each of said straight lines is either horizontal or vertical.

7. The method of claim 6, comprising the additional steps of:

combining any decomposed co-aligned lines into a new set of straight lines;

creating a rectangular table perimeter;

cutting said rectangular table perimeter into smaller rectangles using said straight lines whereby a plurality of cells are created; and

eliminating said group of quads if any text quad is not circumscribed by a rectangular cell.

8. A method of recreating a table in a destination document from a visual display file using quads to encompass textual and graphics elements, comprising the steps of:

combining all groups of closely spaced quads into a polygon frame;

eliminating all frames from further potential table consideration that contain graphic decomposed lines oriented in other than vertical and horizontal;

replacing co-aligned lines, resulting from decomposition of graphics, with a single line;

recreating a table using decomposed replaced and unaltered lines to create cells; and

eliminating frames where textual quads are outside table created cells.

9. A method of converting information in a visual display file using quads to a word processing software compatible type document, comprising the steps of:

combining all groups of quads spaced within a first predetermined distance of one another and comprising at least one of textual symbols and textual associated symbols into an expandable polygon frame;

combining all groups of quads in each frame that are spaced within a second predetermined distance, which second predetermined distance is less than said first predetermined distance, into words; and

converting each of the frames into a word processing type output document.

10. A method of converting table related information, in a visual display source file using quads, to a spreadsheet importable type document, comprising the steps of:

combining all quads of the table related information in the visual display file into a frame;

oversizing all graphic quads in the frame;

decomposing all oversized graphic quads into straight lines;

creating a rectangle of substantially the same size as the table in the visual display document;

replacing co-aligned lines of said straight lines, resulting from decomposition of graphics, with a single line;

recreating a table using decomposed, replaced and unaltered lines to create cells;

inserting textual quads within created cells occupying substantially the same relative space in the created rectangle as it did in the table of the visually displayed source file; and

generating a mined document describing the composition of the created rectangle and the placement of text therein.

11. A method of converting information in a range of visual display file textual matter, comprising multiple paragraphs of text, using quads into separate paragraphs of data in a mined document, comprising the steps of:

combining all groups of quads spaced within a first predetermined distance of one another and comprising at least one of textual symbols and textual associated symbols into an expandable polygon frame;

combining all groups of quads in each frame that are spaced within a second predetermined distance, which second predetermined distance is less than said first predetermined distance, into words; and

converting each of the frames into a separate paragraph in a mined output document.

12. Apparatus for extracting data from a source document wherein words, portions of words and graphic elements are circumscribed by polygons, hereinafter referred to as quads, comprising:

means for oversizing each word and graphic quad in a selected range of the document;

means for assembling quads that overlap into frames;

means for decomposing all graphical elements within each frame into straight lines; and

means for assembling components of any frame into a table when examination of the decomposed elements in a frame indicates a predetermined arrangement of table cell defining straight lines.

13. Apparatus as claimed in claim 12, comprising in addition:

means for retaining relative placement data of each decomposed line as obtained from the source document;

means for combining any co-aligned straight lines after decomposition in a frame ascertained to comprise a table;

means for creating a rectangular table boundary in a frame ascertained to comprise a table;

means for splitting any previously formed rectangle determined to be cut by any selected co-aligned line inserted into said rectangular table boundary to create at least two new rectangles; and

means for generating a mined document from which data may be displayed.

14. Apparatus for mining data from a visually displayable source document having quads defining textual and graphics elements, comprising:

means for combining quads, having less than a predetermined separation, into frames;

means for generating textual paragraphs, in an output document, from frames detected to contain only textual and textual associated symbols; and

means for generating tables, in said output document, from frames detected to contain only vertically and horizontally oriented straight lines enclosing textual symbols.

15. Apparatus for eliminating a group of quads in a source document, each quad of which is relatively situated within a predetermined distance of each other, as potentially defining a table of data, comprising:

means for decomposing all graphic elements in said group of quads into straight lines;

means for determining the orientation of each of said straight lines; and

means for eliminating said group of quads where any of said straight lines is oriented other than horizontal and vertical.

16. Apparatus for establishing that a group of quads in a source document, each of which is relatively situated within a predetermined distance of each other, potentially defines a table of data, comprising:

means for decomposing all graphic elements in said group of quads into straight lines; and

means for determining that the orientation of each of said straight lines is either horizontal or vertical.

17. Apparatus as claimed in claim 16, comprising in addition:

means for combining any decomposed co-aligned lines into a new set of straight lines;

means for creating a rectangular table perimeter;

means for cutting said rectangular table perimeter into smaller rectangles using said straight lines whereby a plurality of cells are created; and

means for eliminating said group of quads as a potential table if any text quad is not circumscribed by a rectangular cell.

18. Apparatus for recreating a table in a destination document from a visual display file using quads to encompass textual and graphics elements, comprising:

means for combining all groups of closely spaced quads into a polygon frame;

means for eliminating all frames from further potential table consideration that contain graphic decomposed lines oriented in other than vertical and horizontal;

means for replacing any set of co-aligned lines, resulting from decomposition of graphics, with a single line;

means for recreating a table using decomposed replaced and unaltered lines to create cells; and

means for eliminating frames as table candidates where textual quads are outside table created cells.

19. Apparatus for converting information in a visual display file using quads to a word processing software compatible type document, comprising:

means for combining all groups of quads spaced within a first predetermined distance of one another and comprising at least one of textual symbols and textual associated symbols into an expandable polygon frame;

means for combining all groups of quads in each frame that are spaced within a second predetermined distance, which second predetermined distance is less than said first predetermined distance, into words; and

means for converting each of the frames into a word processing type output document.

20. Apparatus for converting table related information, in a visual display source file using quads, to a spreadsheet importable type document, comprising:

means for combining all quads of the table related information in the visual display file into a frame;

means for oversizing all graphic quads in the frame;

means for decomposing all oversized graphic quads into straight lines;

means for creating a rectangle of substantially the same size as the table in the visual display document;

means for replacing co-aligned lines of said straight lines, resulting from decomposition of graphics, with a single line;

means for recreating a table using decomposed, replaced and unaltered lines to create cells;

means for inserting textual quads within created cells occupying the same relative space in the created rectangle as it did in the table of the visually displayed source file; and

means for generating a mined document describing the composition of the created rectangle and the placement of text therein.

21. Apparatus for converting information in a range of visual display file textual matter, comprising multiple paragraphs of text, using quads into separate paragraphs of data in a mined document, comprising:

means for combining all groups of quads spaced within a first predetermined distance of one another and comprising at least one of textual symbols and textual associated symbols into an expandable polygon frame;

means for combining all groups of quads in each frame that are spaced within a second predetermined distance, which second predetermined distance is less than said first predetermined distance, into words; and

means for converting each of the frames into a separate paragraph in a mined output document.

22. A method of recreating a table in a destination document from a visual display source file using quads to encompass textual and graphics elements, comprising the steps of:

replacing, in a frame, co-aligned lines, resulting from decomposition of graphics, with a single line;

recreating a table using decomposed replaced and unaltered lines to create cells;

inserting textual quads into cells corresponding to the table location of the textual quad in the source file; and

eliminating rectangles of substantially less than the size of the smallest cell containing a textual quad.

23. A method of positioning textual quads in a column of cells at least some of which contain multiple columns of data in a single column of cells, comprising the steps of:

checking every cell of a source document derived table frame for cells having segregated sets of textual symbol quads;

assigning virtual cut-lines to textual quads that are aligned with textual quads in prior similar size cells in a given column; and

using the virtual cut-lines in aligning textual matter in an output document.

24. Apparatus for recreating a table in a destination document from a visual display source file using quads to encompass textual and graphics elements, comprising:

means for replacing, in a frame, co-aligned lines, resulting from decomposition of graphics, with a single line;

means for recreating a table using decomposed replaced and unaltered lines to create cells;

means for inserting textual quads into cells corresponding to the table location of the textual quad in the source file; and

means for eliminating rectangles of substantially less than the size of the smallest cell containing a textual quad.

25. Apparatus for positioning textual quads in a column of cells at least some of which contain multiple columns of data in a single column of cells, comprising:

means for checking every cell of a source document derived table frame for cells having segregated sets of textual symbol quads;

means for assigning virtual cut-lines to textual quads that are aligned with textual quads in prior similar size cells in a given column; and

means for using the virtual cut-lines in visually aligning textual matter in an output document.